Centre for Research on Peace and Development (CRPD)

# An Assessment of Interviewer Error in the Afrobarometer Project

**Leila Demarest**

CRPD Working Paper No. 53

July 2017

Centre for Research on Peace and Development (CRPD)
KU Leuven
Parkstraat 45, box 3602, 3000 Leuven, Belgium
Phone: +32 16 32 32 50; Fax: +32 16 32 30 88; http://www.kuleuven.be/crpd

# An Assessment of Interviewer Error in the Afrobarometer Project

**Leila Demarest**

## Abstract

This paper focuses on data quality within the Afrobarometer project, which conducts surveys in a wide number of African countries to investigate citizens' beliefs about democracy and political institutions, as well as gain information on their political behaviour. An empirical assessment of interviewer variance is conducted for several key items of the Afrobarometer surveys, for 12 countries and across 3 survey rounds. Interviewer variances are estimated by making use of two-level linear and logistic models, as well as cross-classified models to distinguish between interviewer and area or sampling unit effects. The analyses indicate that interviewer variances are substantial in Afrobarometer data and require further attention. Several recommendations are made to improve Afrobarometer data quality and direct further research.

## Authors

**Leila Demarest** is PhD Fellow at the Centre for Research on Peace and Development (CRPD), University of Leuven (email: Leila.Demarest@kuleuven.be)

## Acknowledgements

The author would like to thank Geert Loosveldt, Line Kuppens, and Maarten Schroyens for valuable comments on earlier drafts of this paper. The majority of the research for this paper was conducted while the author was a PhD Fellow of the Research Foundation Flanders (FWO).

## 1. Introduction

The Afrobarometer project aims to question African citizens on their beliefs about democracy and political institutions, as well as gain information on their political behaviour. The project was founded by Michael Bratton (Michigan State University), Robert Mattes (Institute for Democracy in South Africa), and Emmanuel Gyimah-Boadi (Center for Democratic Development Ghana) in 1999. Currently the Afrobarometer runs as an African-led academic research network. The first survey round covered 12 Sub-Saharan countries. The sixth round was completed in 2015 for over 35 countries, including some countries in North-Africa. Afrobarometer data have been used for an increasing number of academic publications (e.g. Bratton et al., 2010; Bratton, 2013; Eifert et al., 2010), as well as more than 150 working papers within the project. All data are freely available on the project's website.[1]

Given the expansion of the collection and use of Afrobarometer data over time, it is somewhat surprising that only a limited number of studies have focused on data quality within the project. This is in stark contrast with Western-focused projects such as the European Social Survey (ESS), for which a rich methodological literature is available.[2] It is also remarkable as survey quality in developing countries can be particularly challenging due to, for example, the lack of an up-to-date sampling frame, lack of geographical maps, limited access to respondents because of poor road networks; administrative, cultural, and language barriers; different understandings of concepts and survey questions etc. (Bulmer & Warwick, 1983; UN, 2005). These challenges render in-depth documentation on Afrobarometer survey data quality all the more relevant.

This paper is intended as an important step to fill the gap in the methodological literature on Afrobarometer data. Empirically, I focus on interviewer-related effects, specifically interviewer variability in survey data. The use of interviewers as a source of survey error has also led to concern in Western contexts, but it is possible that interviewer-related errors are from a higher level in African settings. Indeed, interviewers often conduct interviews with poorly educated or illiterate respondents, which possibly increases their impact on how questions are understood and/or answered. Furthermore, African countries are typically characterized by ethnic and religious diversity with different cultural identities potentially affecting interviewer-respondent interactions. Finally, interviewers can play crucial roles in the sampling process, especially when the interviewer performs the commonly used random walk method and is responsible for randomly selecting individuals from a given household. This is also the main procedure for Afrobarometer surveys. By investigating interviewer-

[1] www.afrobarometer.org
[2] www.europeansocialsurvey.org

related errors in the Afrobarometer, the paper therefore also aims to contribute to the literature on survey errors in developing countries.

In the following section I discuss the role of the interviewer as a possible source of error in survey designs. I also summarize empirical findings in the literature with regard to interviewer-related error in Western and in African settings. Section 3 describes the role of the interviewer in the Afrobarometer design. Section 4 assesses the magnitude of interviewer variability for key variables in the Afrobarometer surveys. The data used stem from the 12 original Afrobarometer countries in Rounds 3, 4, and 5. I make use of two-level linear and logistic models, as well as cross-classified models to distinguish between interviewer and area (sampling unit) effects. I also conduct several further robustness checks. Section 5 concludes and draws lessons for the Afrobarometer project and future research.

## 2. Interviewer-Related Error in Survey Designs

### 2.1. Interviewer bias

Within the Total Survey Error framework the role of the interviewer is mainly situated at the level of measurement (e.g. Biemer & Lyberg, 2003; De Leeuw et al., 2008; Groves et al., 2004). As with survey errors in general, interviewers can cause bias and variability in the data (Groves et al., 2004, pp. 269-301; Loosveldt, 2008). Interviewer biases are systematic effects of interviewers on respondent answers. The presence of the interviewer can engender social desirability bias, for example, leading the respondent to misreport certain attitudes (e.g. racism) or behaviour (e.g. drug use). This form of bias can be investigated by comparing interviewer and self-administered survey modes. In this paper, I do not focus on this form of interviewer error, however.

Another form of interviewer bias that can arise is related to observable traits of the interviewers (Groves et al., 2004, pp. 269-301; Loosveldt, 2008). It is for example possible that interviewer characteristics such as age, gender, race etc. influence respondents' answers to survey questions. Whether the characteristics of interviewers cause bias of the survey estimate is dependent on the survey design. When only interviewers with the same observable traits are recruited, for example women, estimates can be biased if all respondents underreport certain behaviours in the presence of women. Yet estimates can also be biased if male and female respondents respond differently to the gender of the interviewer and men and women's estimates are compared. The effects of interviewer traits on respondent answers can be measured as fixed effects and interacted with respondent characteristics.

A large number of studies has investigated how interviewers' observable traits such as race, ethnicity, and gender influence respondent's answers (Davis et al., 2009; West and Blom, 2016). In general these studies indicate that interviewer characteristics can influence survey responses, but in particular for survey questions related to these observable traits. However, it is more difficult to draw strong conclusions on the effects of interviewer gender. While most studies focus on Western contexts, evidence from African settings confirms these general findings. In a study by Adida et al. (2015), which is based on Afrobarometer data, the authors analyse the effect of the ethnicity of the interviewer on responses, in particular when interacted with the ethnicity of the respondent. They find that interviewer coethnicity (whether the interviewer has the same ethnicity as the respondent) significantly affects survey responses, in particular for items related to ethnicity. Furthermore, as in Western studies, the effect of interviewer gender is not always as expected. Studies have shown that the effect of interviewer gender can differ by respondents' gender, but also by survey question and setting, which has led to criticism on the common practice of matching interviewer and respondent gender, especially in health surveys, based on untested assumptions (Bignami-Van Assche et al., 2010; McCombie and Anarfi, 2002).

While interviewer bias related to observable characteristics can be measured by testing the effect of interviewer characteristics on respondents' answers, these forms of error can also be reflected in interviewer variability estimates. Indeed, recruited interviewers often have different traits, which in turn have variable effects on respondents' answers to survey questions. Interviewer variability measures are discussed below.

### 2.2. Interviewer variance

Interviewer variability refers to non-systematic differences in the way interviewers perform their tasks (Groves et al., 2004, pp. 269-301; Loosveldt, 2008). Some interviewers make mistakes when reading out questions, skip certain questions or change orders, incorrectly write down survey answers etc. These forms of measurement error can cause additional variance and imprecision in the data. As mentioned, variability can also arise from observable interviewer traits.

Interviewer-related variance can be measured by making use of multilevel statistical models (Kish, 1962). A multilevel model with respondents on the first level, clustered by interviewers on the second level, allows for an examination of the total variation in respondents' answers which can be associated with the interviewer by calculating the intra-cluster (or -class) correlation (ICC). It also reflects the degree of commonality between respondents interviewed by one interviewer. Preferably, the intra-cluster correlation equals zero, indicating that interviewers have no variable effects on respondents.

There is an important caveat when calculating and comparing intra-interviewer correlations, however, namely the assumption that interviewers interview comparable groups of respondents (Groves et al., 2004, pp. 269-301; Loosveldt, 2008). If interviewers overlap with primary sampling units (PSU) for example, the correspondence between respondents can be due to the fact that they originate from the same locality and not because they share the same interviewer. This is important as in many surveys, respondents are indeed not randomly assigned to interviewers. Instead, interviewers often conduct interviews in particular areas to reduce travel costs. Investigating intra-interviewer correlations therefore requires the use of additional control variables (Hox, 1994) or the use of interpenetrated designs (Bailar, 1983; West and Blom, 2017). In interpenetrated designs more than one interviewer conducts interviews in a specific PSU. Interviewer and area effects can be disentangled by making use of cross-classified multilevel models nesting respondents both in interviewers and in sampling units.

Several studies have investigated interviewer-related intra-cluster correlations in specific survey projects, which also allows for some measure of comparison when evaluating Afrobarometer data. Groves (2004, p. 364), for example, notes several studies which show ICCs between 0,01 (1%) and 0,02 (2%) for face-to-face surveys. This leads him to conclude that ICCs below 0,02 are most common. O'Muircheartaigh and Campanelli (1998) review previous studies on intra-interviewer correlations and find values between -0,296 and 0,216 with both minima and maxima from the same study. Most correlations are below 0,1. In their own analyses of 820 variables in a British Household Panel Survey, the authors find ICCs between -0,02 and 0,18 with the large majority under 0,05. They do not find differences in the size or significance of the ICCs between factual and attitudinal items. Furthermore, they find that interviewer effects are of a similar level than the effects of clustering within the Primary Sampling Unit. Schnell and Kreuter (2005) find somewhat different results in a survey on crime perceptions in Germany. They find that interviewer effects are larger than sampling clustering effects, but analyze interviewer effects in terms of design effects. The authors also find that sensitive items produce higher interviewer effects than non-sensitive items, as well as that non-factual items produce higher interviewer effects than factual items. Most recently, Buellens & Loosveldt (2016) find substantial average ICCs for 48 survey items across 6 European Social Survey (ESS) Rounds. Average ICCs can go up to 0,28. Some countries in particular are at a high-risk for interviewer effects.

With regard to interviewer variance in African surveys, a comprehensive study on interviewer effects in two longitudinal household surveys in Kenya and Malawi by Bignami-Van Assche et al. (2010) can serve as an important guideline. For questions on respondents' background and wealth, the ICC tends to fall between 0,01 and 0,07. However, for questions concerning gender relations and AIDS, ICCs are higher and can reach up to 0,25 for questions related to AIDS. They also find that ICCs vary across research sites with generally lower interviewer effects in Malawi than Kenya.

**2.3. Interviewers and nonresponse error**

Although the role of the interviewer is primarily associated with measurement error, there is also a potential relationship between the interviewer and errors of representation, specifically non-response error (Smith, 2011). Interviewers question respondents which can lead to measurement error, but they are often also responsible for the implementation of sampling protocols and for contacting respondents and achieving their cooperation. Poor interviewers, for example, can make more mistakes in the interviewing process (measurement error), and get a lower response rate (nonresponse error). Selective nonresponse can be related to the interviewer's method of approach or an observable trait leading to certain respondents refusing to cooperate.

A substantial strand of the empirical literature on interviewer errors has focused on how interviewers can influence survey nonresponse (West & Blom, 2017). Moorman et al. (1999), for example, find that cooperation rates are higher when the race of the interviewer is matched with the race of the respondent. O'Muircheartaigh and Campanelli (1999) investigate interviewer ICCs for household noncontacts and refusals in a follow-up to their 1998 study discussed in the previous section. They find that interviewer variance is higher than sampling cluster variance for survey nonresponse. Their findings suggest that interviewers who are good at establishing contact are also good at establishing cooperation. This is also supported by research of Pickery & Loosveldt (2002). They make use of the first wave of a Belgian election survey and find that both refusal and noncontact are clustered by interviewer. Furthermore, interviewers who obtain more refusals are also more likely to report noncontact.

Interestingly, in a study by West & Olson (2010), the authors pose the question of how much of interviewer variance is related to nonresponse error instead of measurement error. The authors make use of a dataset on divorce which includes information of respondents as well as nonrespondents as the sample was drawn from county divorce certificates. They calculate interviewer ICCs for the full sample of nonrespondents and respondents, as well as for the respondents-only and compare these measures. Based on their findings, the authors argue that large interviewer ICCs can be due to nonresponse error and differences in the type of respondents interviewers can convince to cooperate, rather than measurement error.

Unfortunately, not many studies seem to focus on interviewer-related nonresponse error in African contexts. However, an interesting study by Kriel & Risenga, (2014) reveals that interviewers can cause error in the sampling process by the way they interpret sampling instructions. Focus groups with interviewers who were involved in a national household survey in South Africa revealed that interviewers often found it difficult to implement the provided definition of a household in the field and therefore operationalized it in different ways on the ground. Weinreb (2006) focuses on differences between locally-recruited interviewers (insiders) and externally recruited interviewers (outsiders) in a longitudinal survey in rural Kenya. He finds that insider interviewers achieve higher cooperation rates, which can be connected to issues of local trust.

The foregoing sections demonstrate that in both Western and African settings interviewer-related errors pose concerns for the quality of survey data projects. This can also apply to the Afrobarometer project, a large cross-national survey project on African citizens' political attitudes and behaviour. In the following section, I discuss the role of the interviewer in the Afrobarometer survey design.

## 3. Afrobarometer Methodology[3]

The role of the interviewer in the Afrobarometer design relates to the face-to-face implementation of the questionnaire as well as the sampling process. The questionnaire is administered in a Paper-Assisted Personal Interview (PAPI). To minimize measurement error related to the interviewer, national partners organize interviewer training sessions. In principle, the interviewer has no translator role. Per participating country a questionnaire translation is made for every language group that is likely to constitute at least 5% of the sample by relying on blind back-translations. However, pre-translated questionnaires do not always preclude on-the-spot-translations in the field as noted by Weinreb and Sana (2009).

The target population of the Afrobarometer surveys are all citizens within the country older than 18. The sample design is a clustered, stratified, multi-stage, area probability sample. None of the countries surveyed by the Afrobarometer can make use of a sampling frame, a list of all members of the target population from which respondents can be randomly drawn. Respondents are generally selected by a random walk method. Several steps are undertaken to acquire a nationally representative sample. The standard sample size for Afrobarometer surveys is 1200 or 2400 cases. 2400 cases are recommended for large, heterogeneous countries. The sampling design is foreseen to lead to an average margin of sampling error of no more than plus or minus 2.8 percent at a 95 percent confidence level.

[3] Information in this section draws from the website and the Afrobarometer Round 6 Survey Manual (Afrobarometer, 2014). No specific manuals are available for Rounds 3, 4, and 5, which are empirically investigated here, but much of the information seems applicable to these rounds as well based on the website.

The sample is stratified according to the first administrative unit of each country and urban/rural residence. Stratification occurs proportionate to population size and information on population sizes is acquired via national censuses. Clustering occurs at the smallest geographic unit for which reliable population data are available. In most countries, these primary sampling units are the census enumeration areas. In general, the Afrobarometer clusters 8 interviews per Primary Sampling Unit (PSU).

Once the PSU is selected, households are chosen based on a list of all households in the enumeration area if such a list is available. Or, more commonly, a starting point for the interviewer walks is randomly selected based on a map. A team of 4 interviewers and 1 supervisor is assigned to a PSU. Hence, the Afrobarometer effectively makes use of an interpenetrated design (Bailar, 1983; West and Blom, 2017). From the random walk starting point, the interviewers conduct the walk in opposite directions. Interviewers use an interval of 5 households for the first interview, followed by an interval of 10 households for the second interview. Once the household is selected, the interviewer list all its members and randomly selects individual respondents using a card drawing method. As the Afrobarometer project aims for an equal proportion of men and women in the sample, interviewers alternate between interviewing a male and a female respondent. In case of nonresponse, households, not individuals within the household, are substituted. Nonresponse occurs when respondents refuse to participate in the survey or when contact could not be established with the household or the individual respondent. Ineligible households include respondents who are not citizens, are deaf or do not speak a survey language, do not fit the gender quota, or do not have an adult member. In case of non-contact, interviewers make one more call later in the day at the household after the first contact attempt. When a respondent refuses, the interviewer replaces the household by continuing the walking pattern. In case of non-contact after two attempts, the interviewer substitutes the household with the very next household on the road. Before leaving the primary sampling area, the field supervisor is instructed to randomly select one of the eight households in which an interview was conducted and check whether the random walk method was implemented correctly.[4]

Within the Afrobarometer survey design, interviewers clearly play crucial roles. They are responsible for the sampling of households and individuals, establishing contact and achieving cooperation, and administering questionnaires in face-to-face interviews without computer aides. Interviewer-related errors in Afrobarometer data can therefore be substantial. In the following section, I investigate interviewer variance in several key items of the Afrobarometer surveys. I first discuss the data and variables analysed, and then conduct

[4] The supervisor back-check variable is included in the datasets for Rounds 3 and 4, but not 5. In Rounds 3 and 4, most countries achieve the guidelines of 12,5% back-checks, except for Lesotho in Round 3, and Botswana and Malawi in Round 4.

both two-level and cross-classified multilevel analyses to assess the magnitude of interviewer-ICCs. I also conduct several additional robustness checks.

## 4. Empirical Analyses

### 4.1. Data and variables

For the analyses of interviewer variance I make use of data from Rounds 3, 4, and 5 for the 12 original Afrobarometer countries: Botswana, Ghana, Lesotho, Malawi, Mali, Namibia, Nigeria, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe. The estimates are hence based on countries which have an established and similar level of experience in implementing the survey. The use of multiple rounds also allows an examination in trends over time.[5]

Although the 12 countries have been implementing the survey since 1999, the national Afrobarometer partners have very different practices in their use of interviewers. Table 1 shows the divergences between countries in the number of interviewers hired, the mean, minimum, and maximum interviews conducted, as well as the average workload (sample size/number of interviewers). Most national Afrobarometer partners seem to follow particular guidelines in terms of the number of interviewers hired. Lesotho and Namibia, for example, always use only a limited number of interviewers, while Nigeria and South-Africa use many. Some changes can occur over time. Nigeria in Round 5 uses about half of the interviewers in Rounds 3 and 4. Furthermore, while Uganda and Zimbabwe roughly double the amount of interviewers when the sample size goes from 1200 to 2400 (Round 5), Ghana and Tanzania maintain the number of interviewers, but increase their workload. Furthermore, the low minima of interviews conducted per interviewer are noticeable. Many interviewers conduct a lot of interviews, yet some only one or a few. While there is no documentation on this issue, this can indicate possible problems with the ways interviewers performed.

[5] The Afrobarometer datasets are freely available online. These data contain interviewer identity codes, but not PSU codes. The PSU data are available on request, yet only subsets of the data are provided. This is an additional explanation for the sample of Afrobarometer surveys used here.

**Table 1: Use of interviewers by Country and Round**

| COUNTRY[a] | ROUND 3 | | | | | | ROUND 4 | | | | | | ROUND 5 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | #[b] | Mean[c] | SD[d] | Min[e] | Max[f] | N[g] | # | Mean | SD | Min | Max | N | # | Mean | SD | Min | Max | N |
| **BOT** | 24 | 50 | 1,051 | 42 | 57 | 1200 | 25 | 48 | 1,995 | 5 | 56 | 1200 | 32 | 37,5 | 0,641 | 34 | 46 | 1200 |
| **GH** | 44 | 27,205 | 0,582 | 19 | 36 | 1197 | 48 | 25 | 1,045 | 2 | 35 | 1200 | 50 | 48 | 1,769 | 28 | 66 | 2400 |
| **LES** | 16 | 82,929 | 9,283 | 1 | 123 | 1161 | 14 | 85,714 | 8,046 | 11 | 102 | 1200 | 17 | 70,412 | 4,546 | 1 | 83 | 1197 |
| **MLW** | 16 | 75 | 1,144 | 66 | 80 | 1200 | 20 | 60 | 5,823 | 3 | 80 | 1200 | 33 | 72,939 | 4,691 | 4 | 89 | 2407 |
| **MLI** | 28 | 44,429 | 3,21 | 1 | 69 | 1244 | 28 | 44 | 3,607 | 3 | 72 | 1232 | 25 | 48 | 2,17 | 4 | 63 | 1200 |
| **NAM** | 16 | 75 | 0,585 | 72 | 78 | 1200 | 17 | 70,588 | 3,83 | 10 | 76 | 1200 | 17 | 70,588 | 4,999 | 1 | 86 | 1200 |
| **NGR** | 133 | 22,363 | 10,775 | 4 | 64 | 2363 | 143 | 16,252 | 0,674 | 3 | 38 | 2324 | 64 | 37,5 | 1,926 | 18 | 79 | 2400 |
| **SAF** | 136 | 17,647 | 1,388 | 1 | 66 | 2400 | 163 | 14,724 | 0,498 | 2 | 32 | 2400 | 104 | 23,067 | 1,595 | 1 | 70 | 2399 |
| **TAN** | 26 | 40,154 | 3,072 | 2 | 66 | 1304 | 20 | 60,4 | 1,317 | 46 | 69 | 1208 | 27 | 88,889 | 5,002 | 6 | 102 | 2400 |
| **UG** | 32 | 75 | 0,624 | 70 | 80 | 2400 | 66 | 36,833 | 2,003 | 2 | 68 | 2431 | 71 | 33,803 | 2,65 | 1 | 94 | 2400 |
| **ZAM** | 43 | 27,907 | 1,346 | 2 | 43 | 1200 | 37 | 32,432 | 2,402 | 1 | 54 | 1200 | 34 | 35,294 | 1,604 | 10 | 52 | 1200 |
| **ZIM** | 20 | 52,4 | 2,226 | 42 | 69 | 1048 | 28 | 42,857 | 3,423 | 1 | 58 | 1200 | 50 | 48 | 1,455 | 1 | 56 | 2400 |

a. Botswana (BOT), Ghana (GH), Lesotho (LES), Malawi (MLW), Mali (MLI), Namibia (NAM), Nigeria (NGR), South Africa (SAF), Tanzania (TAN), Uganda (UG), Zambia (ZAM), Zimbabwe (ZIM).
b. Number of interviewers.
c. Mean interviews per interviewer
d. Standard deviation
e. Minimum number of interviews
f. Maximum number of interviews
g. Sample size

To investigate the level of interviewer variance in Afrobarometer data, I focus on 8 key variables (see Table 2). These have been purposefully selected as responses to these items may be influenced differently by interviewers. For factual items such as Age, Lack of Food[6], and Newspaper Reading, negligible interviewer effects are expected (e.g. Schnell & Kreuter, 2005). For non-factual questions such as Living Conditions, Women in Politics, Democracy, National Identity, and Party Affiliation, higher interviewer effects are expected. It is important to note that the Democracy variable is one of the key variables of the initial Afrobarometer project (Bratton et al., 2010; Bratton, 2013), the National Identity question has been used in a number of academic studies (Langer et al. 2016; Eifert et al., 2010[7]), and the Party Affiliation question has been used in studies on African party systems (Elischer, 2013; Riedl, 2015). The Women in Politics question has mainly been selected because of a hypothesized sensitivity to interviewer gender.

All items are included in Rounds 3, 4, and 5, except for the 'Women in Politics' item, which is only included in Round 5, and the 'National Identity' item, which was not included in Zimbabwe's Round 3 survey. The Age variable is continuous, all other items are binomial, or categorical questions transformed to binomial ones. Such transformations are also used by O'Muircheartaigh and Campanelli (1998, p.65). Here, they were done to avoid response categories with low frequencies and create two categories with the most equal number of respondents possible.

**Table 2: Variables analysed[a]**

| Variable | Question wording | Response scale | Transformed response scale |
|---|---|---|---|
| **Age** | How old are you? | Continuous | / |
| **Living conditions** | In general, how do you rate your living conditions compared to those of other [nationals]? | 1=Much worse, 2=Worse, 3=Same, 4=Better, 5=Much better | 0= Much Worse, Worse, or Same<br>1= Better or Much Better |
| **Lack of Food** | Over the past year, how often, if ever, have you or anyone in your family gone without: Enough food to eat? | 0=Never, 1=Just once or twice, 2=Several times, 3=Many times, 4=Always | 0= at least once<br>1= Never |
| **Newspaper** | How often do you get news from the following sources: | 0=Never, 1=Less than once a month, 2=A few | 0= at least once |

[6] Whether the Lack of Food item is strictly factual can be debated, however, as respondents might attempt to influence interviewer's perceptions of their deprivation.
[7] Eifert et al. (2010) use an older form of the question used in Rounds 1 and 2.

| | | | |
|---|---|---|---|
| | Newspapers? | times a month, 3=A few times a week, 4=Every day | 1= Never |
| **Women in Politics** | Which of the following statements is closest to your view?<br>Statement 1: Men make better political leaders than women, and should be elected rather than women.<br>Statement 2: Women should have the same chance of being elected to political office as men. | 1=Agree very strongly with Statement 1, 2=Agree with Statement 1, 3=Agree with Statement 2, 4=Agree very strongly with Statement 2, 5=Agree with neither | 0= Agree with 1<br>1= Agree with 2<br>Neither= set to missing |
| **Democracy** | Which of these three statements is closest to your own opinion?<br>Statement 1: Democracy is preferable to any other kind of government.<br>Statement 2: In some circumstances, a non-democratic government can be preferable.<br>Statement 3: For someone like me, it doesn't matter what kind of government we have | 1=Statement 3: Doesn't matter, 2=Statement 2: Sometimes non-democratic preferable, 3=Statement 1: Democracy preferable | 0= statement 2 & 3<br>1 = statement 1 |
| **National Identity** | Let us suppose that you had to choose between being a [National] and being a ________ [R's Ethnic Group]. Which of the following best expresses your feelings? | 1=I feel only (R's ethnic group), 2=I feel more (R's ethnic group) than [National], 3=I feel equally [National] and (R's ethnic group), 4=I feel more [National] than (R's ethnic group), 5=I feel only [National] | 0= more ethnic or same<br>1= more national |
| **Party affiliation** | Do you feel close to any particular political party? | 0=No, (not close to any party), 1=Yes, (feels close to a party) | |

a. Refusals, missing, and don't know responses are set to missing.

### 4.2. Methodology

To empirically investigate interviewer variance, I make use of multilevel models and calculate the interviewer-related ICCs (Kish, 1962). As seen in Section 3, the Afrobarometer protocol makes use of 4 interviewers per PSU, which indicates that interviewer variance will not be fully confounded with variance caused by the sampling clusters. Nonetheless, interviewers are still sent to particular regions which can introduce some regional effects in interviewer-ICCs. On the other hand, as they work in teams of 4, interviewers can share certain methods, implying that interviewer effects due to commonalities between team members, can be confounded with PSU effects. Hence, I calculate interviewer and PSU-ICCs separately, and then estimate cross-classified models for each of the 8 variables. The cross-classified models control for the most precise area-level variable, the PSU, in estimating the interviewer-ICCs. These ICCs can hence be fully attributed to the interviewers and not area effects.

For each variable, models (1), (2), and (3) are estimated. While equations (1) and (2) refer to two-level models —with 'INT' clustering by interviewer and 'PSU' clustering by primary sampling unit—, equation (3) reflects the cross-classified model (following the notation of Rabash and Brown, 2001). For binomial dependents the logit transformation is applied. All calculations were conducted in Stata (version 12.1). For the continuous models the xtmixed function was used, for the logistic models the xtmelogit command was used. The logistic cross-classified models were calculated by making use of the Laplace approximation (Rabe-Hesketh and Skrondal, 2012). For all logistic multilevel models the within-variance is fixed to $\pi^2/3 = 3{,}29$, the standard method for estimating these models (Snijders and Bosker, 1999).

(1) $y_{iINT} = \beta_0 + u_{INT} + e_{iINT}$

(2) $y_{iPSU} = \beta_0 + u_{PSU} + e_{iPSU}$

(3) $y_{i(INT,PSU)} = \beta_0 + u_{INT} + u_{PSU} + e_{i(INT,PSU)}$

The ICCs are calculated as:

(1, 2) ρ= between-variance/ (between + within-variance)

(3) ρ= between-interviewer variance/ (between-interviewer variance + between-PSU variance + within variance)

### 4.3. Results

I first look at average ICCs across countries and rounds, and then turn to the item-specific ICCs. Table 3 summarizes the analysis results per country and round.[8] I report the average ICC when clustered by interviewer ($ICC_{INT}$) and by PSU ($ICC_{PSU}$) in the two-level models, and the ICCs based on the cross-classified models ($ICC_{INTcc}$ and $ICC_{PSUcc}$).

[8] All syntax as well as the ICC results for all models are available upon request.

**Table 3: Summary of analysis results**

| COUNTRY | Round 3 | | | | Round 4 | | | | Round 5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $ICC_{INT}$ | $ICC_{PSU}$ | $ICC_{INTcc}$ | $ICC_{PSUcc}$ | $ICC_{INT}$ | $ICC_{PSU}$ | $ICC_{INTcc}$ | $ICC_{PSUcc}$ | $ICC_{INT}$ | $ICC_{PSU}$ | $ICC_{INTcc}$ | $ICC_{PSUcc}$ |
| **BOT** | 0,196 | 0,085 | 0,194 | 0,074 | 0,175 | 0,095 | 0,174 | 0,051 | 0,117 | 0,074 | 0,113 | 0,048 |
| **GH** | 0,159 | 0,103 | 0,151 | 0,069 | 0,186 | 0,129 | 0,171 | 0,055 | 0,264 | 0,176 | 0,256 | 0,048 |
| **LES** | 0,152 | 0,062 | 0,152 | 0,053 | 0,138 | 0,092 | 0,135 | 0,072 | 0,122 | 0,082 | 0,121 | 0,056 |
| **MLW** | 0,097 | 0,081 | 0,096 | 0,060 | 0,105 | 0,117 | 0,100 | 0,096 | 0,143 | 0,104 | 0,139 | 0,052 |
| **MLI** | 0,219 | 0,183 | 0,209 | 0,060 | 0,222 | 0,139 | 0,207 | 0,07 | 0,189 | 0,117 | 0,183 | 0,061 |
| **NAM** | 0,305 | 0,124 | 0,302 | 0,064 | 0,158 | 0,122 | 0,157 | 0,084 | 0,185 | 0,143 | 0,186 | 0,106 |
| **NGR** | 0,061 | 0,280 | 0,057 | 0,118 | 0,268 | 0,26 | 0,230 | 0,115 | 0,201 | 0,129 | 0,195 | 0,061 |
| **SAF** | 0,278 | 0,263 | 0,255 | 0,097 | 0,332 | 0,337 | 0,304 | 0,095 | 0,318 | 0,139 | 0,309 | 0,047 |
| **TAN** | 0,132 | 0,136 | 0,126 | 0,090 | 0,095 | 0,112 | 0,091 | 0,079 | 0,245 | 0,171 | 0,24 | 0,094 |
| **UG** | 0,190 | 0,130 | 0,187 | 0,060 | 0,255 | 0,162 | 0,247 | 0,064 | 0,287 | 0,18 | 0,276 | 0,072 |
| **ZAM** | 0,195 | 0,157 | 0,179 | 0,103 | 0,163 | 0,148 | 0,149 | 0,084 | 0,136 | 0,129 | 0,128 | 0,091 |
| **ZIM** | 0,167 | 0,174 | 0,159 | 0,124 | 0,194 | 0,154 | 0,184 | 0,086 | 0,225 | 0,152 | 0,221 | 0,061 |

Note: ICCs are calculated based on 8 variables in Round 5: Age, Living Conditions, Lack of Food, Newspaper Reading, Women in Politics, Democracy, National Identity, and Party Affiliation. The Women in Politics item is not included in Rounds 4 and 3. The National Identity item is not included in Zimbabwe in Round 3.

It can be indicative to first compare Interviewer- with PSU-ICCs as the relative magnitudes of interviewer and sampling clustering has been debated in the literature (O'Muircheartaigh and Campanelli, 1998; Schnell and Kreuter, 2005). Table 3 shows that Interviewer-ICCs tend to be higher than PSU ICCs; only 6 average PSU-ICCs are higher than the Interviewer-ICCs. The Interviewer-ICCs in the cross-classified models control for PSU variance, but generally do not differ substantially from the ICCs in the two-level models. Differences in the averages lie between -0,001 and 0,038. When controlling for interviewer variance, all average PSU-ICCs become smaller. With the exception of Nigeria in Round 3, all cross-classified PSU-ICCs are smaller than the Interviewer-ICCs. Results correspond to findings by Schnell & Kreuter (2005), who find larger variances attributed to interviewers than to geographical clustering in a German survey.[9]

Average ICCs are found to be quite high in Table 3. Looking at the cross-classified Interviewer-ICCs, it appears that most are higher than 0,1 with a maximum of 0,309 (South-Africa in Round 5) and a minimum of 0,057 (Nigeria in Round 3). Interestingly, although there are differences between countries, it is not the case that some countries clearly perform better than others. This is in contrast with findings from the European Social Survey (Buellens and Loosveldt, 2016). While Nigeria has the lowest average Interviewer-ICCs in Round 3, for example, ICCs are substantially higher in Rounds 4 and 5. For some countries average Interviewer-ICCs increase over time, but they also decline for others. Hence, there does not appear to be a clear learning effect with national partners becoming more experienced in implementing the survey with less errors.[10]

I now investigate item-specific Interviewer-ICCs. Figure 1 demonstrates the variation in Interviewer-ICCs per item, across all countries and rounds, based on the cross-classified models. For the Age, Living Conditions, Lack of Food, Newspaper Reading, Democracy, and Party Affiliation the boxplots are based on 36 estimates. For the Women in Politics item 12 estimates are represented in the boxplot (Round 5 only). For the National Identity item there are 35 observations as the item was not included in Zimbabwe's Round 3 survey.

[9] In cross-classified models differences in the between-variance can also arise from differences in the relative size of clusters (Leckie and Bell, 2013). In general, clusters with more units have lower variances. As interviewer clusters have more units than PSU units, it can be assumed that higher interviewer variances than PSU variances in the cross-classified models reflect real underlying issues with interviewer effects on the data.

[10] It is possible that changes occur in the national partner implementing the survey, but while the current partners are advertised on the website, I could not retrace which partner were involved in each round.

**Figure 1: Interviewer-ICCs by item (R3-R5)**

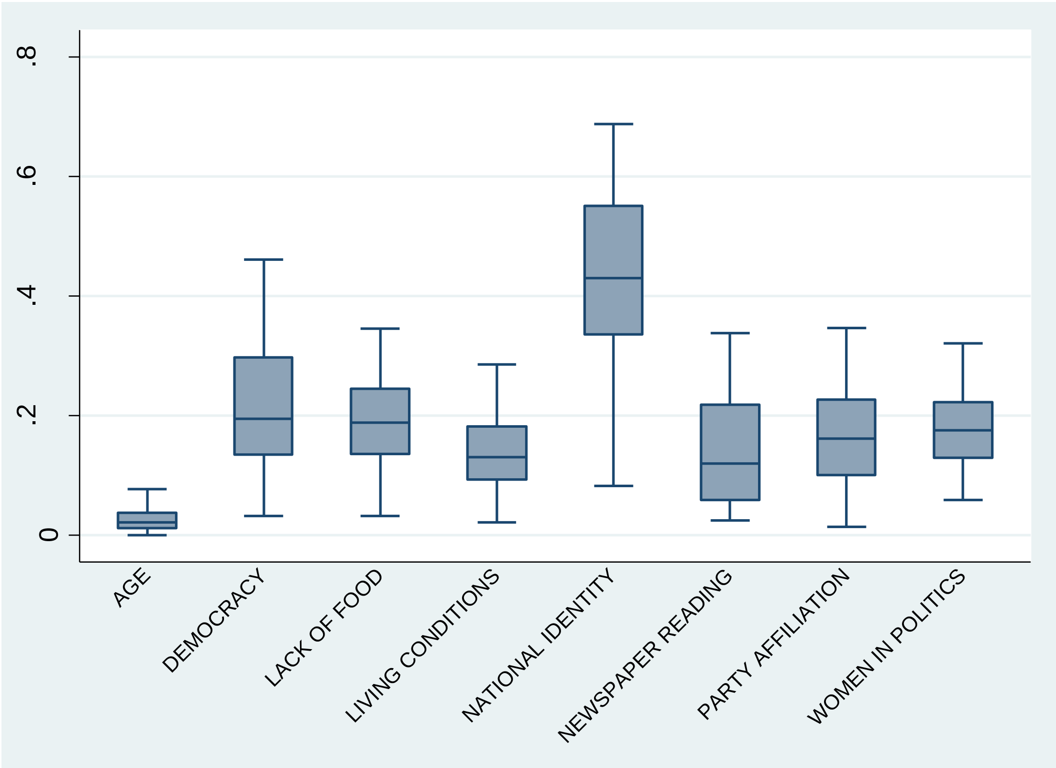



Figure 1 first of all shows that Interviewer-ICCs for the Age variable are generally low, which is as expected. Nonetheless, in South-Africa's Round 4 survey, about 7,5% of variance in the Age question can be explained by the interviewer. In Section 2, it was discussed that previous Western- and African-focused studies generally find intra-interviewer correlations lower than 0,1; maxima cited in the literature were around 0,25. Clearly, Afrobarometer interviewer errors are substantially higher. However, Buellens and Loosveldt (2016) report average ICCs in the ESS up to 0,28, which indicates that for some items ICCs are higher. Besides the Age variable, all variables have averages higher than 0,1. There is also no strong evidence that factual questions such a Lack of Food and Newspaper Reading tend to have lower ICCs than the opinion items. The National Identity variable has the most remarkable results with a minimum of 0,08 for Nigeria (Round 3) and a maximum of 0,69 for South Africa (Round 5). The average lies around 0,43.[11] These results can be related to the findings of Adida et al. (2015), who emphasize the importance of respondent-interviewer coethnicity for survey questions pertaining to ethnicity. The authors acquired additional data on Afrobarometer interviewer ethnicity for this study, however.

[11] Standard errors for the National Identity between variances are also high.

### 4.4. Robustness checks

To check the validity of the above results, I now test whether they hold when using a different statistical programme and estimation technique. This is done specifically for the multilevel logistic models which can be more difficult to estimate than linear models. I also investigate the effect of adding control variables. Finally, I check for several ordinal categorical variables how results are affected by treating them as continuous and estimating the ICCs from linear models. I make use of Round 5 data for these checks.

The logistic model results for equations (1), (2), and (3) are retested in MLwiN (version 2.10). For the two-level logistic models, second-order PQL and MCMC estimation are used. Cross-classified models were recalculated with MCMC (Browne, 2012, pp. 215-230). Table 4 summarizes how the estimates for interviewer-ICCs differ from the Stata estimates used in the previous section. The table provides the mean deviation from the Stata estimate, as well as the largest deviation found for the 7 variables under investigation (the continuous Age variable is excluded). The results demonstrate that estimation technique does play an important role for the size of interviewer ICCs. Second-order PQL tends to estimate lower between-variances, while MCMC estimates larger ones.[12] The differences, in particular for MCMC estimation, are not negligible. The largest differences are generally found for the National Identity variable. The results do support the use of the Stata estimates above as they can be considered relatively conservative.

[12] In several MCMC models not all parameters were estimated well.

**Table 4: Summary of differences between MLwiN and Stata estimates**

| | **Stata $ICC_{INT}$ - PQL2 $ICC_{INT}$** | | **Stata $ICC_{INT}$ - MCMC $ICC_{INT}$** | | **Stata $ICC_{INTcc}$ - MCMC $ICC_{INTcc}$** | |
|---|---|---|---|---|---|---|
| | Mean | Max | Mean | Max | Mean | Max |
| **BOT** | 0,003 | 0,009 | -0,014 | -0,028 | -0,017 | -0,036 |
| **GH** | 0,006 | 0,023 | -0,015 | -0,020 | -0,039 | -0,138 |
| **LES** | -0,002 | 0,001 | -0,019 | -0,050 | -0,046 | -0,198 |
| **MLW** | 0,001 | 0,003 | -0,017 | -0,025 | -0,037 | -0,116 |
| **MLI** | 0,002 | 0,006 | -0,025 | -0,036 | -0,039 | -0,068 |
| **NAM** | -0,001 | 0,001 | -0,037 | -0,055 | -0,071 | -0,034 |
| **NGR** | 0,003 | 0,007 | -0,010 | -0,015 | -0,015 | -0,028 |
| **SAF** | 0,015 | 0,052 | -0,011 | -0,018 | -0,025 | -0,067 |
| **TAN** | 0,001 | 0,005 | -0,026 | -0,037 | -0,052 | -0,107 |
| **UG** | 0,008 | 0,026 | -0,013 | -0,018 | -0,019 | -0,043 |
| **ZAM** | 0,004 | 0,024 | -0,012 | -0,035 | -0,021 | -0,064 |
| **ZIM** | 0,002 | 0,006 | -0,014 | -0,017 | -0,021 | -0,031 |

Next, I test how Interviewer-ICCs are affected by adding fixed effects to the models. I first estimate the Interviewer-ICCs ($ICC_{INT|Pred}$) for the 2-level models after controlling for interviewer characteristics (age, gender, and education) and respondents characteristics (age gender, education).[13] To verify that no area effects continue to confound ICC estimates in the cross-classified models, I also re-test these by controlling for urban/rural residence, region, and ethnic group ($ICC_{INTcc|Pred}$).[14] All models with fixed effects are estimated in Stata (version 12.1).

Table 5 shows how estimates from the models with predictors deviate from those of the null models. Firstly, adding respondent and interviewer characteristics to the models does generally reduce the Interviewer-ICCs. On average the reduction is negligible, however, and limited to a few percentage points which do not affect the overall magnitude of the ICCs.

[13] The variable for interviewer education was dichotomized per country and creates two categories with the most equal distribution of respondents. It was not included in the models for Mali as all interviewers have the same level of education. Most interviewers have higher education levels. Respondent education was dichotomized with 0= not higher than primary schooling, and 1= higher than primary. Interviewer and respondent age were included as continuous variables. Respondent age is excluded from the models with Age as the dependent variable.

[14] These models do not include respondent and interviewer characteristics. For Nigeria, regional information was clustered into the six geopolitical zones of the country. This variable was used instead of the standard regional variable available in the datasets. The ethnic group variable included in the models is based on the one in the dataset, but 5 categories for the largest groups are created with one rest group for other respondents. For some countries, more categories were created if groups contained more than 100 respondents.

Differences are larger for the National Identity question, yet in Mali, the large difference is driven by the Newspaper Reading item. Interestingly interviewer age, gender, and education hardly had a significant effect on the dependent variables in the models. This is consistent with findings in the literature (e.g. West & Blom, 2017). Indeed, even the effect of interviewer gender on the Women in Politics item, was only significant (and positive) for Lesotho.

**Table 5: Summary of differences between fixed effects and null models**

| | $ICC_{INT}$ - $ICC_{INT\|Pred}$ | | $ICC_{INTcc}$ - $ICC_{INTcc\|Pred}$ | |
|---|---|---|---|---|
| | Mean | Max | Mean | Max |
| **BOT** | 0,015 | 0,048 | -0,005 | -0,014 |
| **GH** | 0,026 | 0,094 | 0,107 | 0,189 |
| **LES** | 0,020 | 0,039 | -0,006 | -0,016 |
| **MLW** | 0,013 | 0,055 | 0,025 | 0,034 |
| **MLI** | 0,043 | 0,148 | 0,055 | 0,222 |
| **NAM** | 0,029 | 0,079 | 0,028 | 0,111 |
| **NGR** | 0,010 | 0,045 | 0,025 | 0,038 |
| **SAF** | 0,005 | 0,030 | 0,016 | 0,042 |
| **TAN** | 0,011 | 0,041 | -0,014 | -0,122 |
| **UG** | 0,007 | 0,033 | 0,036 | 0,088 |
| **ZAM** | 0,018 | 0,039 | 0,025 | 0,053 |
| **ZIM** | 0,016 | 0,033 | 0,049 | 0,092 |

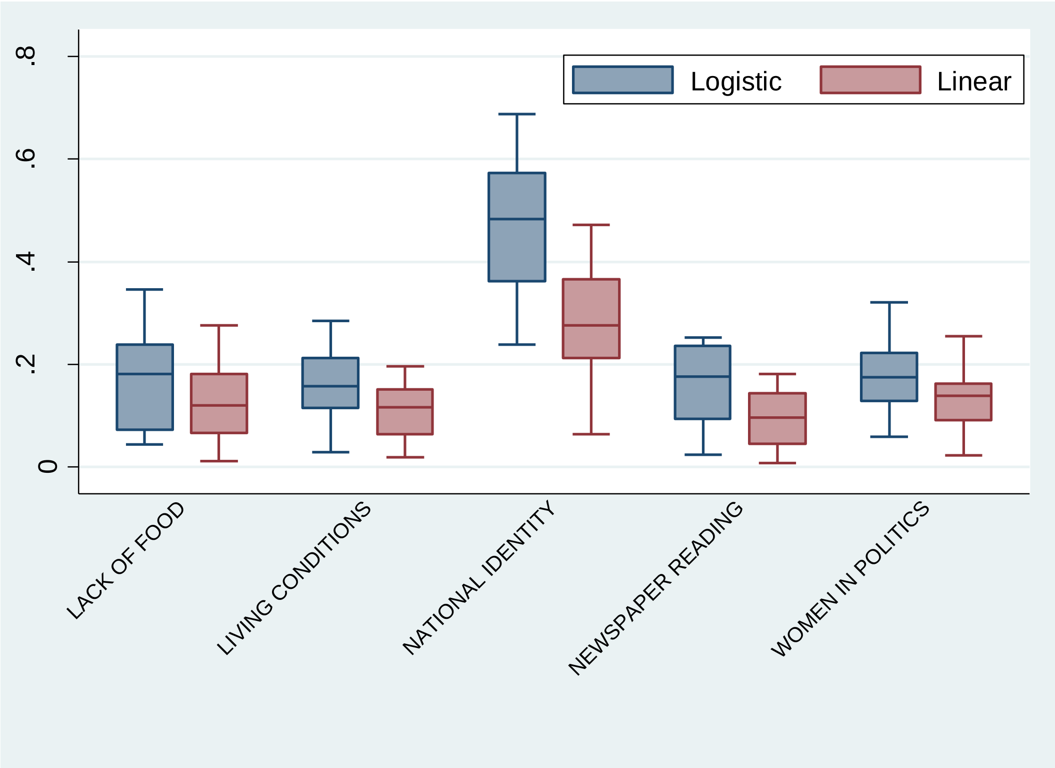
Adding further control variables to the cross-classified models mostly reduces interviewer-ICCs although there are exceptions. On average, the deviations are limited, taking into account the substantial ICCs in the null models.[15] The case of Ghana does stand out, and large reductions occur for nearly all variables. Yet in contrast to the other countries considered here, Ghana's interviewers are fully nested within the regions which can explain these results. For the other countries, the largest deviations tend to occur for the Newspaper Reading item.

Finally, I estimate the magnitude of the Interviewer-ICCs for the Living Conditions, Lack of Food, Newspaper Reading, Women in Politics, and National Identity variables in linear cross-classified models. As can be seen from Table 2, all variables are ordered categorical with a limited number of categories. Figure 2 shows the boxplots for the Interviewer-ICCs.

[15] Removing interviewer clusters with less than 4 respondents from these models also does not affect results (the average deviation equals -0,0005).

The ICCs tend to be lower in the linear versus the logistic models. This can be due to the variable transformation, but also the assumption that the within-variance of the logistic models equals 3,29 (Snijders & Boskers, 1999). Yet even in the linear models, ICCs tend to be substantial as they still have means above 0,10 with high maxima.

**Figure 2: Interviewer-ICCs by item (R5), linear vs logistic cross-classified models**




## 5. Conclusion

The Afrobarometer project conducts public opinion surveys in a wide range of African countries. As cross-national survey projects can be particularly challenging in developing contexts, this is a remarkable endeavor. However, it remains important to critically evaluate the quality of Afrobarometer data and devise ways to reduce survey errors. Even though multiple rounds of Afrobarometer surveys have been conducted, relatively little research has been devoted to this aim. This paper has focused on interviewer variance in Afrobarometer data and has shown that interviewer effects can be substantial and require more attention.

It is first of all important to recognize that high interviewer variance can be related to measurement error, or the effect interviewers have on respondents during the interview, as well as nonresponse error, or differences in the types of respondents who are recruited by the interviewers (West & Olson, 2010). During the interview, interviewers can influence respondents by not following interview protocols, performing on-the-spot translations which lead to different interpretations, or by eliciting certain responses because of observable traits. Ethnicity in particular can have an important effect and could potentially explain high interviewer variances for the National Identity survey item (see also Adida et al., 2015). Contact and response rates reported for Afrobarometer surveys are generally high[16], but random walk methods are not always reliable and interviewers can have considerable leeway in the selection of households and individuals within households. This can lead to underreported nonresponse and biases in the pools of respondents recruited by interviewers.

To assess the causes of high interviewer variance in Afrobarometer data more information on the survey process as well as the interviewers is needed. Switching to Computer-Assisted Personal Interviewing (CAPI) can reduce mistakes during the interview (and data input) process, but can also allow for the systematic collection of data on interview duration and speed per question, the tracking of interviewer walk patterns via GPS coordinates, as well as the time between interviews (e.g. Byass et al., 2008; Hattas & Eloff, 2011). These data allow for a detailed monitoring of the interviewer process and more quality control. Furthermore, the Afrobarometer could collect further data on the interviewers and make these available for researchers (upon signing of a privacy protection agreement). In the framework of interviewer training, for example, interviewers can familiarize themselves with the questionnaire by filling it out themselves. These responses could be stored and later compared with respondents'.

The project could also take a closer look at interviewer hiring practices by national partners and the extent to which recommendations provided (see Afrobarometer, 2014) are followed. Remuneration of interviewers can be a particularly thorny issue as well as their workload. High interviewer variances together with high interviewer workloads lead to substantial design effects and can reduce the effective sample size considerably (Loosveldt, 2008).[17] This implies that survey estimates are less precise than they appear. As seen in Table 1, some countries make use of only a limited number of interviewers, while others use many. For equal Interviewer-ICCs, the design effect is larger in countries where interviewers have higher workloads. Guidelines with regard to interviewer workloads can reduce these design effects.

Given the concerns on data quality raised in the empirical analyses, further research and additional documentation on relevant quality indicators by the Afrobarometer project are needed. The project can draw on the substantial methodological documentation behind the

[16] www.afrobarometer.org
[17] The design effect related to the interviewer (deffINT) can be calculated as: $deffINT = 1+ \rho INT * (\text{mean workload}-1)$, with $\rho$ indicating the ICC.

European Social Survey, for example. While the Afrobarometer provides valuable data on African citizens' political attitudes, additional research is needed to ensure the validity and reliability of substantive research findings. For researchers working with Afrobarometer data, it can be recommended to control whether empirical findings hold when taking into account interviewer effects.

## References

Adida, C.L., Ferree, K.E., Posner, D.N. and Robinson, A.L. (2015). Who's asking? Interviewer coethnicity effects in African survey data. Afrobarometer Working Paper 158.

Afrobarometer (2014). Round 6 Survey Manual. www.afrobarometer.org

Bailar, B.A. (1983). Interpenetrating Subsamples. In S. Kotz and N.L. Johnson (Eds.) Encyclopedia of Statistical Sciences, 4, 197–201. New York: Wiley.

Biemer, P.P. and Lyberg, L.E. (2003). Introduction to survey quality. Hoboken: Wiley.

Bignami-Van Assche, S., Reniers, G. and Weinreb, A. (2003). An Assessment of the KDICP and MDICP Data Quality. Demographic Research, 1(2), 31- 76

Bratton, M., Ed. (2013). Voting and Democratic Citizenship in Africa. Boulder: Lynne Riener.

Bratton, M., Mattes, R. and Gyimah-Boadi, E. (2010). Public opinion, democracy and market reform in Africa. Cambridge: Cambridge University Press.

Browne, W.J. (2012). MCMC estimation in MLwiN. Centre for Multilevel Modelling, University of Bristol.

Buellens, K. and Loosveldt, G. (2016). Interviewer effects in the European Social Survey. Survey Research Methods, 10(2), 103-118

Bulmer, M. and Warwick, D.P., Eds. (1983). Social research in developing countries
surveys and censuses in the Third World. Chichester: Wiley.

Byass, P., Hounton, S., Ouédraogo, M., Somé, H., Diallo, I., Fottrell, E., Emmelin, A. and Meda, N. (2008). Direct data capture using hand-held computers in rural Burkina Faso: experiences, benefits and lessons learnt. *Tropical Medicine & International Health, 13*(1), 25-30

Davis, R.E., Couper, M.P., Janz, N.K., Caldwell, C.H. and Resnicow, K. (2010). Interviewer effects in public health surveys. Health Education Research, 25(1), 14–26

De Leeuw, E.D.,Hox, J.J. and Dilman, D.A. (2008). The Cornerstones of survey research. In E.D. de Leeuw, J.J. Hox and D.A. Dilman (Eds.), International handbook of survey methodology (pp. 1-17). New York: Erlbaum.

Eifert, B., Miguel E. and Posner, D.N. (2010). Political Competition and Ethnic Identification in Africa. American Journal of Political Science, 54(2), 494–510.

Elischer, S. (2013). Political parties in Africa: ethnicity and party formation. New York: Cambridge University Press.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E. and Tourangeau, R. (2004). Survey Methodology. Hoboken: Wiley.

Hattas, M. and Eloff, M. (2011). Secure digital data collection in household surveys – Case study: Statistics South Africa. IEEE AFRICON Conference. [6072142] DOI: 10.1109/AFRCON.2011.6072142

Hox, J.J. (1994). Hierarchical Regression Models for Interviewer and Respondent Effects. Sociological Methods Research, 22(3), 300-318.

Kish, L. (1962). Studies of interviewer variance for attitudinal variables. Journal of the American Statistical Association, 57(297), 92-115

Kriel, A. and Risenga, A. (2014). Breaking the silence: Listening to interviewers when considering sources of non-sampling errors in household survey research in South Africa. South African Review of Sociology, 45(2), 117-136

Langer, A., Stewart, F., Smedts, K. & Demarest. L. (2016). Conceptualising and Measuring Social Cohesion in Africa: Towards a Perceptions-Based Index. Social Indicators Research. First online: 30 January 2016.

Leckie, G. and Bell, A. (2013). Cross-Classified Multilevel Models – MLwiN Practical. LEMMA VLE Module 12. P.12.2.3 ‘Calculating coverage intervals, variance partition coefficients (VPCs) and intraclass correlation coefficients (ICCs)’. http://www.bristol.ac.uk/cmm/learning/course.html

McCombie, S.C. and Anarfi, J.K. (2002). The Influence of Sex of Interviewer on the Results of an AIDS Survey in Ghana. *Human Organization, 61*(1), 51-57

Moorman, P.G., Newman, B., Millikan, R.C., Tse, C.K. and Sandler, D.P. (1999). Participation Rates in a Case-Control Study: The Impact of Age, Race, and Race of Interviewer. Annals of Epidemiology, 9(3), 188-195.

O'Muircheartaigh, C. and Campanelli, P. (1998). The relative impact of interviewer effects and sample design effects. Journal of the Royal Statistical Society: Series A (Statistics in Society), 161(1), 63-77

O'Muircheartaigh, C. and Campanelli, P. (1999). A multilevel exploration of the role of interviewers in survey non-response. Journal of the Royal Statistical Society: Series A (Statistics in Society), 162(3), 437-446.

Pickery, J. and Loosveldt, G. (2002). A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse. Quality and Quantity, 36(4), 427-437.

Rasbash, J. and Browne, W. (2001). Modelling non-hierarchical structures. In A. H. Leyland and H. Goldstein (eds.), Multilevel Modelling of Health Statistics (pp.93-106). Chichester: Wiley

Rabe-Hesketh, S. and Skrondal, A. (2012). Multilevel and longitudinal modeling using Stata. College Station: Stata Press.

Riedl, R.B. (2015). Authoritarian origins of democratic party systems in Africa. Cambridge: Cambridge University Press 2015.

Schnell, R. and Kreuter, F. (2005). Separating interviewer and sampling-point effects. Journal of Official Statistics, 21(3), 389–410

Snijders, T.A.B. and Bosker, R.J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modelling.* London: Sage.

Smith, T.W. (2011). Refining the Total Survey Error Perspective. International Journal of Public Opinion Research, 23(4), 464-480

UN (2005). Household Sample Surveys in Developing and Transition Countries. Department of Economic and Social Affairs, Statistics Division. Studies in Methods, Series F No 96. New York.

Weinreb, A.A. (2006). The Limitations of Stranger-Interviewers in Rural Kenya. American Sociological Review, 71(6), 1014-1039

Weinreb, A.A. & Sana, M. (2009). The Effects of Questionnaire Translation on Demographic Data and Analysis. Population Research and Policy Review, 28(4), 429-454

West, B.T. and Blom, A.G. (2017). Explaining interviewer effects: A research synthesis. Journal of Survey Statistics and Methodology, 5(2), 175-211.

West, B.T. and Olson, K. (2010). How Much of Interviewer Variance is Really Nonresponse Error Variance? Public Opinion Quarterly, 74(5), 1004-1026